SPARSE REPRESENTATIONS FOR LIMITED DATA TOMOGRAPHY

Hstau Y. Liao[*]

Laboratory of Computational Biology and Macromolecular Imaging, HHMI, Wadsworth Center, Albany, NY 12201, USA

Guillermo Sapiro[†]

Dept. of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455, USA

ABSTRACT

In limited data tomography, with applications such as electron microscopy and medical imaging, the scanning views are within an angular range that is often both limited and sparsely sampled. In these situations, standard algorithms produce reconstructions with notorious artifacts. We show in this paper that a sparsity image representation principle, based on learning dictionaries for sparse representations of image patches, leads to significantly improved reconstructions of the unknown density from its limited angle projections. The presentation of the underlying framework is complemented with illustrative results on artificial and real data.

Keywords: Limited angle tomography, sparse representations, regularization

1. INTRODUCTION

Tomography, with applications such as electron microscopy, medical imaging, and industrial non-destructive testing, refers to the recovery of the density distribution inside the body from its given projections. We are primarily interested in the class of tomography which can be modeled by the Radon transform. In *limited data tomography*, data are collected over an angular range that is either limited (due to physical constraints) or sparsely sampled (e.g., due to cost savings or radiation limitations). The use of standard reconstruction algorithms, such as filtered back-projection (FBP) methods, produces reconstructions with notorious artifacts, see Figure 1.

In dealing with the ubiquitous limited angle tomography, several approaches have been tested (e.g., see [1, 2, 3, 4, 5] for more recent ones). In terms of artifacts, methods that apply regularization in the image (density) domain show higher degrees of success. Nevertheless, they normally assume piecewise smoothness of the unknown image and are still vulnerable to the artifacts, unless a "reference image" is used, which is not always available.

For the sparse angular sampling problem, *total variation* (related) methods have been shown to be very promising, mainly when applied to piecewise constant images [1, 6, 7].

[*]Work supported by a postdoctoral fellowship at the IMA, U Minnesota.
[†]Work supported in part by NSF, NGA, ONR, DARPA, NIH, and ARO.

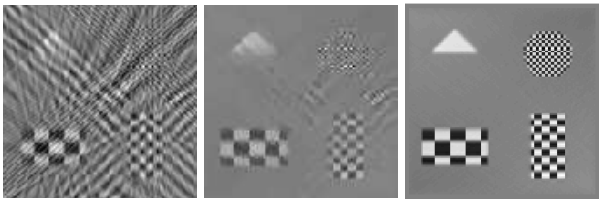

Fig. 1. *Reconstructions of the phantom in Figure 2 from just* 11 *noiseless projections extended over* 120 *degrees. FBP (left), a total variation based method (center), and the results by our proposed method (right).*

It was shown in [6] that if the unknown signal is sparse in some given representation (e.g., in some vector space), then it can be accurately recovered, with high probability, from only a few (random) measurements. An example is the exact recovery of a piecewise constant image like the Shepp-Logan Phantom (which has only a few non-zero gradients) from only a few projections. In practice, images in real applications are seldom piecewise constant, and therefore finding an efficient representation "basis" in which the unknown image is sparse remains an open problem. Here we address this issue from a practical viewpoint, by first considering small image *patches*. We show that by assuming sparsity of the patches with respect to a basis that in turn is being learned (following [8]), we can reconstruct images that cannot be efficiently recovered by these TV-based methods, see Figure 1. We should add that the theoretical results in [6] do not address the important case of the missing wedge, meaning that not only the data is sparsely sampled but also a continuous range of projections is missing (often about 30% of the total range).

2. SPARSITY MODELS IN TOMOGRAPHY

2.1. Sparsity representation of patches

The present work is motivated by the image processing success of the *Sparseland* model for signal recovery problems [9]. For signals from a class $\Gamma \subset \mathbb{R}^N$, this model suggests the existence of a specific redundant dictionary $D \in \mathbb{R}^{N\times K}$ that contains K *atoms*, such that for any signal $x \in \Gamma$, there exists a sparse linear combination of atoms from D that approxi-

mates it well. More formally, this means that $\forall x \in \Gamma,\ \exists\, \alpha \in \mathbb{R}^K,$ such that $x \approx D\alpha$ and $\|\alpha\|_0 \ll N$.

D can be predefined (such as wavelets) or learned (e.g., by the K-SVD algorithm [8]), as in this work. Due to its highly effectiveness for tasks such as image denoising, demosaicing, and inpainting, in particular when the dictionary is learned [9, 10], here we extend this idea to tomographic reconstruction.

To make this framework practical, the Sparseland model, like many other image-domain regularization methods, considers the processing of small overlapping image patches, i.e., such patches are the ones that admit a sparse representation. Assume that the patches are of size $\sqrt{n} \times \sqrt{n}$ (i.e., n pixels in each patch), then the idea is that a patch Mx can be approximated by $D\alpha$, where M is a $n \times N$ binary matrix that extracts the patch from the image x, $D \in \mathbb{R}^{n \times K}$ is the learned dictionary, and $\alpha \in \mathbb{R}^{K \times J}$, with J being the number of patches. The goal then is to learn D such that α is sparse and x is efficiently and accurately reconstructed by the joint sparse representation of all its corresponding overlapping patches. We now briefly present the framework while revisiting the application of this model to image denoising.

2.2. Image denoising model

In [9], the authors considered the classical model for image degradation, $y = x + w$, where $x \in \mathbb{R}^N$ is the clean image, $w \in \mathbb{R}^N$ is assumed to be white Gaussian noise with a *fixed* standard deviation σ (the case of non-uniform σ is dealt with in [10]), and $y \in \mathbb{R}^N$ is the noisy observed image. The energy minimization formulation corresponding to the simultaneous learning of D, computation of α, and restoration of x, is

$$\left\{\widehat{\alpha}, \widehat{D}, \widehat{x}\right\} = \arg\min_{\alpha, D, x} \left\{ \lambda \|x - y\|_2^2 + \sum_{j=1}^{J} \mu_j \|\alpha_j\|_0 + \sum_{j=1}^{J} \|D\alpha_j - M_j x\|_2^2 \right\}, \quad (1)$$

where λ and the μ_j $(j = 1, ..., J)$ are positive, and the columns of $\alpha \in \mathbb{R}^{K \times J}$ are the coefficient vectors α_j $(j = 1, ..., J)$, in a way that the j-th patch $M_j x$ is approximated by $D\alpha_j$. The vector $\widehat{x} \in \mathbb{R}^N$ is an estimate of the true image. The first term in (1) enforces the matching to the data. The second and the last terms provide regularization, considering that the solution has a sparse representation for every overlapping patch over the learned dictionary $\widehat{D}$. Details on the optimization of this variational formulation are presented below.

2.3. Image reconstruction model: The tomography case

For reconstruction in limited angle tomography, we model the measurement $y \in \mathbb{R}^I$ as $y = Rx + w$, where $R \in \mathbb{R}^{I \times N}$ is the (discrete) Radon transform (I projections), and $w \in \mathbb{R}^I$ is the noise, which causes (in general non-uniform) uncertainties in the pixel (or voxel, if in 3D) intensities of the reconstructed image. If the reconstruction process is linear, these uncertainties can be estimated (see, e.g., [11]). This however is not the case, if we attempt to use the regularization in (1), due to the non-linearity introduced by the operation $\|\cdot\|_0$, which affects the (non-deterministic) α. Nevertheless, assuming a deterministic α and uniform noise, we can restore linearity and compute a first order approximation to these uncertainties. It turns out that the largest uncertainties occur mainly at the boundaries within the first few pixels, whereas in the interior they are relatively uniform. As a result, we propose to solve the tomographic reconstruction problem via

$$\left\{\widehat{\alpha}, \widehat{D}, \widehat{x}\right\} = \arg\min_{\alpha, D, x} \left\{ \lambda \|Rx - y\|_2^2 + \sum_{j=1}^{J} \mu_j \|\alpha_j\|_0 + \sum_{j=1}^{J} \|D\alpha_j - M_j x\|_2^2 \right\}, \quad (2)$$

subject to the condition of known boundary within a few pixels, due to the large uncertainties there. Note that while the measurements are in the Radon domain (first term in (2)), the imposed sparsity and learned dictionary are in the image (density) domain.

2.4. Optimization algorithm

The proposed algorithm follows [8], and consists of steps of the K-SVD algorithm (downloaded from www.cs.technion.ac.il/~elad/) for image denoising, alternated with a Radon inversion algorithm. Let $d_1, ..., d_K$ be the columns of D, which are the atoms in the dictionary. Let $\alpha^k \in \mathbb{R}^J$ denote the k-th row of α. The K-SVD algorithm for denoising of gray scale images essentially minimizes the objective function in (1) with respect to the "coordinates" $\alpha_j (j = 1, ..., J)$, $d_k (k = 1, ..., K)$, $\alpha^k (k = 1, ..., K)$, and x. First (*the sparse coding step*), the minimization is carried out with respect to each α_j $(j = 1, ..., J)$ via standard orthogonal matching pursuit (find the best atoms from the given dictionary), and then (*the dictionary update step*) the algorithm optimizes the last term of (1) with respect to d_k and α^k simultaneously (where only the nonzero components of α^k are considered, improve via SVD the atom for all the patches that selected it in the sparse coding step), for each $k = 1, ..., K$. This process is repeated a few times, until D and α have been learned. Finally (*the averaging step*), the objective function is minimized with respect to x (this is a simple weighted average of the overlapping patches). Here we skip the averaging step, optimizing (2) with respect to the image x, which is a simple quadratic optimization solved using Matlab. There is a critical "sparsity threshold" ε in the KSVD algorithm: higher (lower) ε induces fewer (more) atoms to approximate a patch in the sparse coding step. It is determined by the pixel uncertainties that, as discussed above, in the case of tomography are difficult to

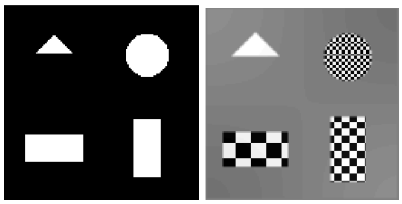

Fig. 2. 128×128 *mathematical phantom with a foreground that has a uniform density (left,* Phantom A*), and checkerboard-like patterns (right, Phantom B).*

compute. Therefore, we set ε as a free parameter. See [10] for more discussion about this parameter.

3. EXPERIMENTAL RESULTS

In the experiments presented below, we used (see (2)) a Lagrange multiplier $\lambda = 1$, size of the patches $n = 25$, size of the redundant dictionary $K = 64$, and known boundary values up to three pixels thick. The initial dictionary was the Discrete Cosine Transform.

3.1. Simulated data

We created two mathematical 2D phantoms. Both phantoms contain four primitives, with constant density in the first one, called *Phantom A*, and checkerboard-like patterns in the second one, called *Phantom B*. See Figure 2. The former is binary with background intensity 0.1 and foreground intensity 1; and the latter has a minimal intensity -1 and maximal intensity 1. On a gray scale, black represents the minimum and white represents the maximum.

The simulated data sets consisted of 11 projections extended uniformly over two thirds of the standard (e.g., in electron tomography) full 180° range. Typically, when standard reconstruction algorithms (such as the FBP) are used, many more (the order of hundreds) projections extending over the full angular range are required. We considered both perfect and noisy projection data. The noise in the measurements was independent Gaussians with unit standard deviation.

For comparisons, we performed the reconstructions using a FBP method, a TV-based method (see Appendix), and the proposed approach. See figures 1 and 3 . Note that, because of the missing angular range, which is in the horizontal direction, edges along this direction are notoriously difficult to recover. We used $\varepsilon = 0.01$ and $\varepsilon = 0.05$, respectively, for the Phantom A and Phantom B. In all cases, the number of iterations was $H = 1,000$.

These preliminary results suggest that our proposed reconstruction method outperforms both qualitatively (e.g., less artifacts and more contrast) and quantitatively a FBP method and a TV-based method. See Table 1, where the estimation error is defined to be $\|\widehat{x} - x_0\|_2$, with $\widehat{x}$ and x_0, respectively, the reconstructed and the true image.

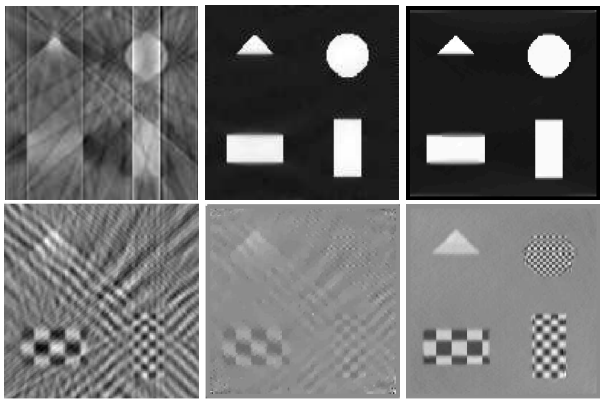

Fig. 3. *Top row: 128×128 reconstructions of the Phantom A from 11 noiseless projections. FBP (left), a TV-based algorithm (center), and the results by our proposed method (right). Bottom row: same order but for the Phantom B and noisy data. Note that TV-based minimization works well with Phantom A (piecewise constant) but not with Phantom B.*

Table 1. *Estimation error .*

Technique	A/B (noiseless)	A/B (noisy)
FBP	20/28	18.6/23
TV minimization	4.3/19	11.1/21.5
Proposed	4.8/10.8	8.7/12.2

3.2. Real data

We performed reconstructions on 221×221 images from dental data produced by the Focus intraoral X-ray source and the Sigma intraoral sensor (Instrumentarium Dental; courtesy of Maaria Rantala, from PaloDex, see [5] for more details). We have available 23 projections, uniformly distributed over approximately the full 180° of a section of the third molar tooth. A typical section has four basic components: surrounding air, a layer of enamel, interior dentin, and pulp (the two dark holes), see Figure 4. To produce satisfactory results, it was necessary only $H = 100$ iterations. To test the effectiveness of our method, we reconstructed from (i) all the 23 projections, (ii) 15 (out of the 23) consecutive projections (leaving then a wedge of uncovered angles), and (iii) 8 (out of the 23) projections approximately uniformly distributed over the full 180°. In the first two cases, ε was lowered linearly from 0.005 to 0.001 in 100 iterations, and in the case (iii), from 0.01 to 0.001. Again, we observe from Figure 4 that our approach delivers reconstructions with less artifacts, without sacrificing the contrast.

4. CONCLUSIONS AND DISCUSSION

In this work, we introduced the framework of learning sparse representations for density reconstruction in limited angle tomography. The reconstruction algorithm aims at minimizing

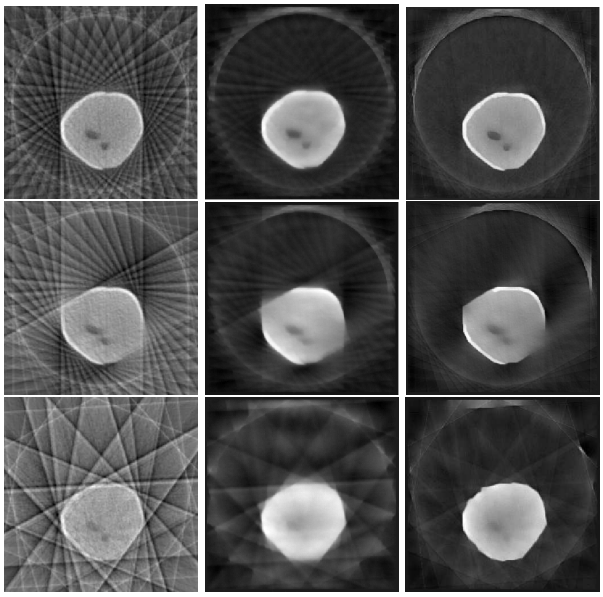

Fig. 4. *Reconstructions of a section of a tooth from* 23 *(top row),* 15 *(middle row) and* 8 *(bottom row) projections, using a FBP method (left column), a total-variation based method (center column), and our proposed method (right column).*

a functional, encouraging a sparse representation of the image patches while keeping the data constraints provided by the available projections. Preliminary experimental results from both simulated and real data demonstrated that the proposed reconstruction method outperforms a FBP and a TV-based method.

As mentioned above, the sparsity-level ε is closely related to the non-uniform uncertainty level, and it was here left as an algorithm parameter. Automatically computing this parameter is the subject of important future research.

Results in the image enhancement literature have shown that significant improvements, leading to state-of-the-art reconstruction, can be achieved by initially considering a dictionary learned from large databases from the same data class [8, 10], which are then adapted to the particular image. We plan to continue our line of research in this direction, and results will be reported elsewhere.

Appendix- A TV-based algorithm: Regarding the implementation of the TV-based method, we found that a simple alternation between TV-minimization and imposing the data constraints, produced better results than solving a constrained TV minimization using, e.g., *l1-magic* http://www.acm.caltech.edu/l1magic/. The results by the l1-magic (using the default parameters and starting with the FBP reconstruction), were very similar to the starting image. The algorithm for all the TV reconstructions in this paper consisted of 3,000 to 4,000 iterations of: (i) one step of gradient descent for the TV minimization with fixed step size, and (ii) five cycles of ART. In general, the reconstructions did not improve considerably after 2,000 iterations and the optimal step size was between 10^{-3} and 10^{-4}.

Acknowledgments: The authors are thankful to Maaria Rantala for the dental data and to Doug Arnold and Julien Mairal for helpful discussions.

5. REFERENCES

[1] I. Aganj, A. Bartesaghi, M. Bognia, H.Y. Liao, G. Sapiro, and S. Subramaniam, "Regularization for inverting the radon transform with wedge consideration," in *Proc. Fourth IEEE Int. Symp. on Biomed. Imag.*, Arlington, VA, 2007, pp. 217–220, IEEE.

[2] A.H. Delaney and Y. Bresler, "Globally convergent edge-preserving regularized reconstruction: an application to limited-angle tomography," *IEEE. Trans. Image Process.*, vol. 7, pp. 204–221, 1998.

[3] V. Kolehmainen, S. Siltanen, S. Järvenpää, J.P. Kaipio, E. Somersalo, M. Lassas, P. Koistinen, and J. Pirttilä, "Statistical inversion for medical X-ray tomography with few radiographs 2: Application to dental radiology," *Phys. Med. Biol.*, vol. 48, pp. 1465–1480, 2003.

[4] H.Y. Liao, "An gradually unmasking method for limited data tomography," in *Proc. Fourth IEEE Int. Symp. on Biomed. Imag.*, Arlington, VA, 2007, pp. 820–823, IEEE.

[5] M. Rantala, S. Vänskä, S. Järvenpää, M. Kalke, M. Lassas, J. Moberg, and S. Siltanen, "Wavelet-based reconstruction for limited angle X-ray tomography," *IEEE Trans. Med. Imag.*, vol. 25, pp. 210–217, 2006.

[6] E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory*, vol. 52, pp. 489–509, 2006.

[7] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, pp. 259–268, 1992.

[8] M. Aharon, M. Elad, and A.M. Bruckstein, "The K-SVD: An algorithm for designing of overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, pp. 4311–4322, 2006.

[9] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, pp. 3736–3745, 2006.

[10] J. Mairal, M. Elad, and G. Sapiro, "Sparse representation for color image restoration," *IEEE Trans. Image Process.*, 2008, to appear.

[11] J.A. Fessler, "Mean and variance of implicitly defined biased estimators (such as penalized maximum likelihood): applications to tomography," *IEEE Trans. Image Process.*, vol. 5, pp. 493–506, 1996.